Biometrika (2015), **102**, 4, *pp.* 981–987
Printed in Great Britain
doi: 10.1093/biomet/asv039
Advance Access publication 15 October 2015

On the validity of the pairs bootstrap for lasso estimators

By L. CAMPONOVO

Faculty of Mathematics and Statistics, University of St. Gallen, Bodanstrasse 6, 9000 St. Gallen, Switzerland

lorenzo.camponovo@unisg.ch

Summary

We study the validity of the pairs bootstrap for lasso estimators in linear regression models with random covariates and heteroscedastic error terms. We show that the naive pairs bootstrap does not provide a valid method for approximating the distribution of the lasso estimator. To overcome this deficiency, we introduce a modified pairs bootstrap procedure and prove its consistency. Finally, we consider the adaptive lasso and show that the modified pairs bootstrap consistently estimates the distribution of the adaptive lasso estimator.

Some key words: Heteroscedastic regression model; Lasso estimator; Pairs bootstrap.

1. Introduction

Consider the linear regression model

$$y_t = \sum_{i=1}^{p} \beta_i x_{t,i} + \epsilon_t \quad (t = 1, \ldots, n), \tag{1}$$

where y_t is the response variable, $x_t = (x_{t,1}, \ldots, x_{t,p})^{\mathrm{T}}$ is a vector of covariates, ϵ_t is the error term, and $\beta = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is the unknown parameter of interest. Throughout the paper, we assume that p is fixed while n is large. Furthermore, we assume that the covariate vector x_t is random and that the error term ϵ_t may be related to x_t.

In the first part of the paper, we focus on the lasso estimator

$$\hat{\beta}_n = \arg\min_u \sum_{t=1}^{n} (y_t - u^{\mathrm{T}} x_t)^2 + \lambda_n \sum_{i=1}^{p} |u_i|, \tag{2}$$

where $\lambda_n > 0$ is a tuning parameter. Since its introduction by Tibshirani (1996), the lasso has been widely used for point estimation and variable selection. In several settings, it may be preferred to alternatives such as the least-squares estimator; see, e.g., Bühlmann & van de Geer (2011).

Knight & Fu (2000) derived the limit distribution of the lasso estimator in linear regression models with nonrandom covariates and homoscedastic error terms. Wagener & Dette (2012) extended the results of Knight & Fu (2000) by considering heteroscedastic error terms. Using similar arguments, we derive the limit distribution of the lasso estimator in linear regression models with random covariates and heteroscedastic error terms. In particular, we show that under some regularity conditions, the law of $T_n = n^{1/2}(\hat{\beta}_n - \beta)$ converges weakly to that of $T = \arg\min_u R(u)$, where $R(u)$ is a random process. As highlighted in Chatterjee & Lahiri (2011), limit distributions of this form are quite complicated, and in practice bootstrap procedures may provide better approximations of the sampling distribution of lasso statistics.

In this setting, a standard procedure is the pairs bootstrap; see, e.g., Freedman (1981). Let $(z_1, \ldots, z_n)$ be the observed data, where $z_t = (y_t, x_t^{\mathrm{T}})^{\mathrm{T}}$. The pairs bootstrap constructs random samples $(z_1^*, \ldots, z_n^*)$ by selecting from $(z_1, \ldots, z_n)$ uniformly with replacement. Let $\hat{\beta}_n^*$ be the solution of (2) based on a bootstrap sample. The pairs bootstrap approximates the sampling distribution of T_n with the conditional distribution of $n^{1/2}(\hat{\beta}_n^* - \hat{\beta}_n)$ given the observations $(z_1, \ldots, z_n)$. In this paper, we show that this approach does not consistently estimate the distribution of T_n, and we identify two different reasons for its failure. First, the regression parameter β satisfies the population moment condition $E\{(y_t - x_t^{\mathrm{T}}\beta)x_t\} = 0$, but $E^*\{(y_t^* - x_t^{*\mathrm{T}}\hat{\beta}_n)x_t^*\} \neq 0$, where E^* denotes the expectation with respect to the distribution of the bootstrap sample conditional on the original sample; therefore, the bootstrap moment condition based on the lasso estimator does not correctly mimic the population moment condition. Second, the lasso estimator does not identify the signs of zero coefficients of the regression parameter with sufficiently high probability; because of this inaccuracy, the bootstrap lasso estimation criterion does not properly penalize zero coefficients of the regression parameter.

To overcome these problems, we introduce a modified pairs bootstrap procedure. First, we recentre the bootstrap statistic with respect to the least-squares estimator $\bar{\beta}_n$ instead of $\hat{\beta}_n$. The least-squares estimator is defined as the solution to (2) with $\lambda_n = 0$, and $\bar{\beta}_n$ satisfies $E^*\{(y_t^* - x_t^{*\mathrm{T}}\bar{\beta}_n)x_t^*\} = 0$. Therefore, the least-squares estimator correctly mimics the population moment condition $E\{(y_t - x_t^{\mathrm{T}}\beta)x_t\} = 0$. Second, we replace the standard lasso estimation criterion with an adjusted bootstrap lasso estimation criterion that properly penalizes zero coefficients of the regression parameter. With these corrections, we show that the modified pairs bootstrap consistently estimates the distribution of T_n.

In § 3, we focus on the adaptive lasso estimator

$$\check{\beta}_n = \arg\min_u \sum_{t=1}^{n}(y_t - u^{\mathrm{T}}x_t)^2 + \lambda_n \sum_{i=1}^{p} \lambda_{n,i}|u_i|, \tag{3}$$

where $\lambda_n > 0$ is a tuning parameter, and $\lambda_{n,i} = |\bar{\beta}_{n,i}|^{-1}$ with $\bar{\beta}_n = (\bar{\beta}_{n,1}, \ldots, \bar{\beta}_{n,p})^{\mathrm{T}}$. Zou (2006) showed that in linear regression models with nonrandom covariates and homoscedastic error terms, the adaptive lasso performs valid model selection and estimates nonzero coefficients of the regression parameter with the same efficiency as the least-squares estimator on the true subset model. Wagener & Dette (2012) extended the results of Zou (2006) by considering heteroscedastic error terms. Using similar arguments, we show that the adaptive lasso also has these properties in our setting. Furthermore, we study the validity of the pairs bootstrap in approximating the sampling distribution of $n^{1/2}(\check{\beta}_n - \beta)$, and show that the modified pairs bootstrap consistently estimates the distribution of the adaptive lasso estimator.

Recently, many authors have proposed inference procedures for lasso estimators. Some rely on resampling methods, such as Chatterjee & Lahiri (2010, 2011, 2013) and Minnier et al. (2011). Another class of procedures considers a desparsification approach, which removes the bias introduced by shrinkage and constructs appropriate approximate inverses of the empirical Gram matrix; see, e.g., Zhang & Zhang (2014) and van de Geer et al. (2014). Lockhart et al. (2014) proposed a covariance statistic for testing the significance of predictors, and in a 2014 unpublished paper J. Taylor, R. Lockhart, R. J. Tibshirani and R. Tibshirani (arXiv: 1401.3889) introduced a new class of test statistics for forward stepwise and least-angle regression that produces exact post-model-selection p-values. Our work is mostly related to that of Chatterjee & Lahiri (2010, 2011), which focuses on the residual bootstrap for homoscedastic linear regression models. Our results supplement their findings by studying the validity of the pairs bootstrap for more general heteroscedastic regression models.

2. PAIRS BOOTSTRAP FOR THE LASSO ESTIMATOR

2·1. *The naive pairs bootstrap*

Before presenting the main results, we introduce following assumption.

Assumption 1. The vectors z_t are independent with $E(\|z_t\|^4) < \infty$, where $\|\cdot\|$ denotes the Euclidean norm. The parameter β minimizes $E\{(y_t - x_t^{\mathrm{T}}\beta)^2\}$. Let $C = E(x_t x_t^{\mathrm{T}})$ be positive definite, and suppose that as $n \to \infty$, the law of $n^{-1/2}\sum_{t=1}^{n} x_t\epsilon_t$ converges weakly to the normal distribution with mean zero and variance matrix Ω.

Assumption 1 is also required in proving the consistency of the pairs bootstrap for the least-squares estimator; see, e.g., Freedman (1981). Moreover, it implies that $E\{(y_t - x_t^{\mathrm{T}}\beta)x_t\} = 0$. In the next lemma we give the limit distribution of T_n.

LEMMA 1. *Suppose that Assumption 1 holds. Furthermore, let $W \sim N(0, \Omega)$ and let $I(\cdot)$ denote the indicator function. If $n^{-1/2}\lambda_n \to \lambda_0 \geqslant 0$, then the law of $T_n = n^{1/2}(\hat\beta_n - \beta)$ converges weakly to that of $T = \arg\min_u R(u)$ as $n \to \infty$, where*

$$R(u) = -2u^{\mathrm{T}}W + u^{\mathrm{T}}Cu + \lambda_0 \sum_{i=1}^{p}\{u_i \,\mathrm{sgn}(\beta_i) I(\beta_i \neq 0) + |u_i| I(\beta_i = 0)\}.$$

When $\lambda_0 = 0$, $T = C^{-1}W \sim N(0, C^{-1}\Omega C^{-1})$, so in this case we obtain the same limit distribution as for the least-squares estimator. On the other hand, when $\lambda_0 > 0$, it may be quite complicated to construct confidence sets for the regression parameter β.

To verify the validity of the pairs bootstrap approximation, we adopt the approach taken in the proof of Lemma 1. First, we show that $T_n^* = n^{1/2}(\hat\beta_n^* - \hat\beta_n)$ minimizes a particular random process $R_n^*(u)$. Then, we compute the limit $R^*(u)$ of $R_n^*(u)$. To this end, we consider the conditional probability given the sample $(z_1, \ldots, z_n)$ and compute the limit $R^*(u)$ by successively conditioning on a sequence of samples, as $n \to \infty$. Finally, we compare $R^*(u)$ with $R(u)$. Consider the process

$$R_n^*(u) = \sum_{t=1}^{n}\left\{(\hat\epsilon_t^* - u^{\mathrm{T}}x_t^* n^{-1/2})^2 - (\hat\epsilon_t^*)^2\right\} + \lambda_n \sum_{i=1}^{p}\left\{|\hat\beta_{n,i} + u_i n^{-1/2}| - |\hat\beta_{n,i}|\right\},$$

where $\hat\epsilon_t^* = y_t^* - x_t^{*\mathrm{T}}\hat\beta_n$. Then we can easily verify that $R_n^*(u)$ is minimized at T_n^*. Consider the first term $R_{n,1}^*(u) = \sum_{t=1}^{n}\{(\hat\epsilon_t^* - u^{\mathrm{T}}x_t^* n^{-1/2})^2 - (\hat\epsilon_t^*)^2\}$. After some algebra we obtain

$$R_{n,1}^*(u) = -2n^{-1/2}\sum_{t=1}^{n} u^{\mathrm{T}}x_t^* x_t^{*\mathrm{T}}(\bar\beta_n - \hat\beta_n) - 2n^{-1/2}\sum_{t=1}^{n} u^{\mathrm{T}}x_t^*\bar\epsilon_t^* + n^{-1}\sum_{t=1}^{n} u^{\mathrm{T}}x_t^* x_t^{*\mathrm{T}}u,$$

where $\bar\epsilon_t^* = y_t^* - x_t^{*\mathrm{T}}\bar\beta_n$. Under Assumption 1, as $n \to \infty$, $n^{-1}\sum_{t=1}^{n} x_t^* x_t^{*\mathrm{T}}$ converges in conditional probability to C, while the conditional law of $n^{-1/2}\sum_{t=1}^{n} x_t^*\bar\epsilon_t^*$ converges weakly to the normal distribution with mean zero and variance matrix Ω; see, e.g., Freedman (1981, Theorem 3.1).

Let $U = n^{1/2}(\hat\beta_n - \beta)$. Then, as in Knight & Fu (2000), we can show that the limit of the second term $R_{n,2}^*(u) = \lambda_n \sum_{i=1}^{p}\{|\hat\beta_{n,i} + u_i n^{-1/2}| - |\hat\beta_{n,i}|\}$ is

$$R_2^*(u) = \lambda_0 \sum_{i=1}^{p}\left\{u_i \,\mathrm{sgn}(\beta_i) I(\beta_i \neq 0) + (|u_i + U_i| - |U_i|) I(\beta_i = 0)\right\}.$$

Furthermore, letting $D = n^{1/2}(\bar\beta_n - \hat\beta_n)$, we can conclude that the limit of $R_n^*(u)$ is

$$R^*(u) = -2u^{\mathrm{T}}CD - 2u^{\mathrm{T}}W + u^{\mathrm{T}}Cu + R_2^*(u).$$

By comparing $R(u)$ with $R^*(u)$, we note two important differences. First, the term $-2u^{\mathrm{T}}CD$ appears only in $R^*(u)$. Second, the penalization of zero coefficients in $R(u)$ and $R^*(u)$ is slightly different. The source of the term $-2u^{\mathrm{T}}CD$ in the definition of $R^*(u)$ is related to the distortion $E^*\{(y_t^* - x_t^{*\mathrm{T}}\hat\beta_n)x_t^*\} \neq 0$. The different penalization of zero coefficients is related instead to the inaccuracy of the lasso in identifying the sign of zero coefficients with sufficiently high probability. This different penalization is also the source of the inconsistency of the naive residual bootstrap for homoscedastic regression models; see, e.g., Chatterjee & Lahiri (2010). Using arguments similar to those in Chatterjee & Lahiri (2010), we can formally prove that the naive pairs bootstrap is also inconsistent.

2·2. *The modified pairs bootstrap*

To overcome the problem of inconsistency of the naive residual bootstrap, Chatterjee & Lahiri (2011) proposed a thresholding procedure. More precisely, in the implementation of the residual bootstrap, they propose replacing $\hat{\beta}_n$ by the modified lasso estimator $\tilde{\beta}_n = (\tilde{\beta}_{n,1}, \ldots, \tilde{\beta}_{n,p})^{\mathrm{T}}$, where $\tilde{\beta}_{n,i} = \hat{\beta}_{n,i} I(|\hat{\beta}_{n,i}| \geqslant a_n)$, with a_n being a sequence of real numbers such that

$$a_n + (n^{-1/2} \log n) a_n^{-1} \to 0 \tag{4}$$

as $n \to \infty$. Examples of sequences that satisfy condition (4) include sequences of the form $a_n = cn^{-\delta}$ for $c \in (0, \infty)$ and $\delta \in (0, 1/2)$. The thresholding has no impact on nonzero coefficients of β for large n. Indeed, if $\beta_i \neq 0$, then $|\hat{\beta}_{n,i}| > a_n$ for large n with high probability. On the other hand, this approach sets the estimates of zero coefficients of β to 0 with high probability. Indeed, if $\beta_i = 0$, then $|\hat{\beta}_{n,i}| < a_n$ for large n with high probability.

Obviously, this thresholding approach cannot be applied to our setting. Nevertheless, in our case it is possible to overcome the distortions of the naive pairs bootstrap. We propose the following corrections. First, to overcome the distortion $E^*\{(y_t^* - x_t^{*\mathrm{T}} \hat{\beta}_n) x_t^*\} \neq 0$, we recentre the bootstrap statistic with respect to $\bar{\beta}_n$ instead of $\hat{\beta}_n$. Indeed, $\bar{\beta}_n$ satisfies $E^*\{(y_t^* - x_t^{*\mathrm{T}} \bar{\beta}_n) x_t^*\} = 0$, which exactly mimics the population moment condition satisfied by the regression parameter. Second, we introduce the modified bootstrap lasso estimator

$$\dot{\beta}_n^* = \arg\min_u \sum_{t=1}^n (y_t^* - u^{\mathrm{T}} x_t^*)^2 + \lambda_n \sum_{i=1}^p \left| u_i - \bar{\beta}_{n,i} I(|\bar{\beta}_{n,i}| \leqslant a_n) \right|, \tag{5}$$

where the sequence a_n satisfies condition (4). In the penalization term in (5), we recentre with respect to $\bar{\beta}_{n,i} I(|\bar{\beta}_{n,i}| \leqslant a_n)$. This recentring has no impact on nonzero coefficients of β for large n. Indeed, if $\beta_i \neq 0$, then $|\bar{\beta}_{n,i}| > a_n$ for large n with high probability, and consequently $u_i - \bar{\beta}_{n,i} I(|\bar{\beta}_{n,i}| \leqslant a_n) = u_i$. On the other hand, if $\beta_i = 0$, then $|\bar{\beta}_{n,i}| \leqslant a_n$ for large n with high probability, and consequently $u_i - \bar{\beta}_{n,i} I(|\bar{\beta}_{n,i}| \leqslant a_n) = u_i - \bar{\beta}_{n,i}$. Therefore, the penalization term in (5) shrinks the bootstrap estimates of zero coefficients of β to the least-squares estimates. This adjustment exactly mimics the standard lasso penalization term that shrinks the estimates of zero coefficients of β to 0. Finally, we approximate the sampling distribution of T_n by the conditional distribution of $n^{1/2}(\dot{\beta}_n^* - \bar{\beta}_n)$ given the observations $(z_1, \ldots, z_n)$. In the next theorem, we establish the validity of the modified pairs bootstrap.

THEOREM 1. *Suppose that Assumption 1 holds. If $n^{-1/2}\lambda_n \to \lambda_0 \geqslant 0$, then the conditional law of $n^{1/2}(\dot{\beta}_n^* - \bar{\beta}_n)$ converges weakly to $T = \arg\min_u R(u)$ as $n \to \infty$.*

Theorem 1 shows that the modified pairs bootstrap provides a valid approximation to the sampling distribution of T_n. Furthermore, by adapting Corollary 3.2 of Chatterjee & Lahiri (2011) to our setting, we can show that the modified pairs bootstrap can be used to construct confidence sets for the regression parameter.

3. PAIRS BOOTSTRAP FOR THE ADAPTIVE LASSO ESTIMATOR

In this section, we analyse the validity of the pairs bootstrap in approximating the sampling distribution of $n^{1/2}(\check{\beta}_n - \beta)$. Before presenting the main results, we introduce some notation. Let $A = \{i : \beta_i \neq 0\}$ and $A^{\mathrm{c}} = \{i : \beta_i = 0\}$. Let $\beta_A = (\beta_{A,1}, \ldots, \beta_{A,q})^{\mathrm{T}}$ and $\beta_{A^{\mathrm{c}}} = (\beta_{A^{\mathrm{c}},1}, \ldots, \beta_{A^{\mathrm{c}},p-q})^{\mathrm{T}}$ denote the subvectors of nonzero and zero coefficients of β, respectively, where $q \leqslant p$. Further, let $\bar{\beta}_{n,A}$ and $\check{\beta}_{n,A}$ denote, respectively, the least-squares and adaptive lasso estimators of β_A; similarly, let $\bar{\beta}_{n,A^{\mathrm{c}}}$ and $\check{\beta}_{n,A^{\mathrm{c}}}$ denote, respectively, the least-squares and adaptive lasso estimators of $\beta_{A^{\mathrm{c}}}$. Finally, let V_A be the asymptotic variance of the least-squares estimator on the true subset model, and let $\check{A}_n = \{i : \check{\beta}_{n,i} \neq 0\}$. In the next lemma, we present the asymptotic properties of the adaptive lasso.

LEMMA 2. *Suppose that Assumption 1 holds. If $\lambda_n \to +\infty$ and $n^{-1/2}\lambda_n \to 0$, then:*

(i) *as $n \to \infty$, $n^{1/2}\check{\beta}_{n,A^{\mathrm{c}}}$ converges in probability to 0;*

(ii) *as $n \to \infty$, the law of $n^{1/2}(\check{\beta}_{n,A} - \beta_A)$ converges weakly to the normal distribution with mean zero and variance V_A;*
(iii) $\lim_{n\to\infty} \mathrm{pr}(\check{A}_n = A) = 1$.

Lemma 2 says that the adaptive lasso performs correct model selection and estimates nonzero coefficients of the regression parameter with the same efficiency as the least-squares estimator on the true subset model.

Unlike the least-squares and lasso estimators, the adaptive lasso sets the estimates of zero coefficients of β to 0 with high probability. However, in this case one also has $E^*\{(y_t^* - x_t^{*\mathrm{T}}\check{\beta}_n)x_t^*\} \neq 0$. Therefore, to overcome this distortion, it is necessary to recentre the pairs bootstrap statistic with respect to $\bar{\beta}_n$ instead of $\check{\beta}_n$. Further, because of this recentring, it is also necessary to modify the bootstrap adaptive lasso estimation criterion. To this end, we introduce the modified bootstrap adaptive lasso estimator

$$\ddot{\beta}_n^* = \arg\min_u \sum_{t=1}^n (y_t^* - u^{\mathrm{T}} x_t^*)^2 + \lambda_n \sum_{i=1}^p \lambda_{n,i}^* \big| u_i - \bar{\beta}_{n,i} I(\check{\beta}_{n,i} = 0) \big|, \tag{6}$$

where $\lambda_{n,i}^* = |\bar{\beta}_{n,i}^*|^{-1}$, with $\bar{\beta}_n^*$ being the bootstrap least-squares estimator. In the penalization term in (6), we recentre with respect to $\bar{\beta}_{n,i} I(\check{\beta}_{n,i} = 0)$. By adopting this correction, the penalization term sets the bootstrap estimates of zero coefficients of β to the least-squares estimates with high probability. Finally, we approximate the sampling distribution of $n^{1/2}(\check{\beta}_n - \beta)$ by the conditional distribution of $n^{1/2}(\ddot{\beta}_n^* - \bar{\beta}_n)$ given the observations $(z_1, \ldots, z_n)$. Let $\ddot{\beta}_{n,A}^*$ and $\ddot{\beta}_{n,A^c}^*$ denote the modified bootstrap adaptive lasso estimators of β_A and β_{A^c}, respectively. In the next theorem, we assert the validity of the modified pairs bootstrap.

THEOREM 2. *Suppose that Assumption 1 holds. If $\lambda_n \to +\infty$ and $n^{-1/2}\lambda_n \to 0$, then as $n \to \infty$:*

(i) *$n^{1/2}(\ddot{\beta}_{n,A^c}^* - \bar{\beta}_{n,A^c})$ converges in conditional probability to 0;*
(ii) *the conditional law of $n^{1/2}(\ddot{\beta}_{n,A}^* - \bar{\beta}_{n,A})$ converges weakly to the normal distribution with mean zero and variance V_A.*

Theorem 2 states that the modified pairs bootstrap provides a valid approximation to the sampling distribution of $n^{1/2}(\check{\beta}_n - \beta)$. Moreover, by extending Corollary 4.2 of Chatterjee & Lahiri (2011) to our setting, we can show that the modified pairs bootstrap can be used in the construction of confidence sets for the regression parameter.

4. MONTE CARLO SIMULATIONS

We use similar simulation designs as in Minnier et al. (2011). In particular, we study the accuracy of inference based on the least-squares estimator and normal approximation, the naive pairs bootstrap for the lasso and adaptive lasso estimators, and the modified pairs bootstrap for the lasso and adaptive lasso estimators. For the lasso estimators, we select the tuning parameters $\lambda_n \in [0, 3n^{1/2}]$ and $a_n \in (0, 0{\cdot}2)$ according to the data-driven method introduced in Remark 2. The number of random samples is $N = 5000$, and the number of bootstrap replications is $B = 799$.

In the first example, we consider model (1) with $n = 100$ or 150 and $p = 12$. The true β contains three large coefficients, $\beta_1 = \beta_2 = \beta_3 = 1$, three moderate coefficients, $\beta_4 = \beta_5 = \beta_6 = 0{\cdot}5$, and six noise coefficients, $\beta_7 = \cdots = \beta_{12} = 0$. For the covariates and error terms, we assume $x_{t,i} \sim N(0, 1)$ and $\epsilon_t \sim N(0, \sigma_j)$ $(j = 1, 2)$, with $\sigma_1 = 1$ and $\sigma_2 = p^{-1}\sum_{i=1}^p x_{t,i}^2$. In Table 1, we report empirical coverage levels and mean lengths of symmetric 95% confidence intervals for large, moderate and noise coefficients. For nonzero coefficients, the modified pairs bootstrap provides empirical coverages very close to 0·95. In contrast, the empirical coverages using the naive pairs bootstrap are quite far from the nominal coverage probability. For zero coefficients, the empirical coverages of normal approximation with the least-squares estimator are slightly smaller than 0·95. On the other hand, bootstrap methods with the adaptive lasso provide shorter confidence intervals with coverage converging to 1. Indeed, Lemma 2 shows that the adaptive lasso

Table 1. *Empirical coverages* ($\times 100$) *and lengths of confidence intervals* (*in parentheses*)

	$\beta_i = 1$		$\beta_i = 0{\cdot}5$		$\beta_i = 0$	
$n = 100$	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2
Norm LS	91·9 (0·39)	91·7 (0·40)	92·2 (0·39)	92·1 (0·40)	92·1 (0·39)	91·7 (0·40)
Boot L	63·5 (0·47)	63·4 (0·49)	61·3 (0·42)	61·3 (0·44)	99·5 (0·16)	99·6 (0·16)
Boot AL	82·8 (0·52)	81·9 (0·54)	59·1 (0·44)	57·1 (0·46)	99·9 (0·09)	99·9 (0·10)
M Boot L	94·7 (0·45)	94·3 (0·46)	94·9 (0·45)	94·2 (0·46)	97·6 (0·34)	97·4 (0·35)
M Boot AL	94·6 (0·44)	94·6 (0·45)	95·9 (0·48)	95·7 (0·50)	99·1 (0·21)	99·0 (0·22)
	$\beta_i = 1$		$\beta_i = 0{\cdot}5$		$\beta_i = 0$	
$n = 150$	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2
Norm LS	92·9 (0·32)	92·7 (0·33)	93·2 (0·32)	92·5 (0·33)	93·1 (0·32)	92·6 (0·33)
Boot L	61·8 (0·36)	60·9 (0·38)	62·5 (0·35)	61·1 (0·37)	99·8 (0·12)	99·8 (0·12)
Boot AL	86·4 (0·37)	84·6 (0·40)	78·8 (0·37)	74·3 (0·39)	99·9 (0·07)	99·9 (0·08)
M Boot L	94·7 (0·36)	94·4 (0·37)	95·1 (0·36)	94·3 (0·37)	97·6 (0·27)	97·2 (0·28)
M Boot AL	94·7 (0·34)	94·3 (0·36)	96·1 (0·38)	96·0 (0·40)	99·6 (0·13)	99·5 (0·15)

Norm LS, least-squares estimator and normal approximation; Boot L, naive pairs bootstrap for the lasso estimator; Boot AL, naive pairs bootstrap for the adaptive lasso estimator; M Boot L, modified pairs bootstrap for the lasso estimator; M Boot AL, modified pairs bootstrap for the adaptive lasso estimator.

Table 2. *Empirical rejection frequencies* ($\times$ 100) *of* $H_0 : \beta_{10} = 0$

$n = 150$	$\beta_{10} = 0$	$\beta_{10} = n^{-1/2}$	$\beta_{10} = 2n^{-1/2}$	$\beta_{10} = 3n^{-1/2}$	$\beta_{10} = 4n^{-1/2}$
Norm LS	6·9	26·9	65·3	90·4	98·6
M Boot L	3·2	24·4	62·8	88·1	98·4
M Boot AL	1·8	17·3	50·8	83·5	96·8

Norm LS, least-squares estimator and normal approximation; M Boot L, modified pairs bootstrap for the lasso estimator; M Boot AL, modified pairs bootstrap for the adaptive lasso estimator.

estimates of zero coefficients collapse to 0 asymptotically. Therefore, in this case the coverage of confidence intervals should converge to 1 as n increases; see e.g., Minnier et al. (2011) for similar empirical findings. Similar arguments also explain the large coverages of bootstrap confidence intervals with the lasso estimator.

In the second example, we consider model (1) with $n = 150$ and $p = 10$. For the covariates and error terms we assume $x_{t,i} \sim N(0, 1)$ and $\epsilon_t \sim N(0, 1)$. The true β contains three large coefficients, $\beta_1 = \beta_2 = \beta_3 = 1$, three moderate coefficients, $\beta_4 = \beta_5 = \beta_6 = 0{\cdot}5$, three noise coefficients, $\beta_7 = \beta_8 = \beta_9 = 0$, and $\beta_{10} = c$ with $c \in [0, 4n^{-1/2}]$. In Table 2 we report the empirical rejection frequencies of the null hypothesis $H_0 : \beta_{10} = 0$ versus the alternative $H_1 : \beta_{10} > 0$, for $c \in [0, 4n^{-1/2}]$ and significance level 0·05. For these parameter selections, the conventional t-statistic ranges from 0 to 4. When $c = 0$, the rejection frequencies using normal approximation with the least-squares estimator are slightly higher than the significance level. On the other hand, in line with the previous example, the modified pairs bootstrap used with lasso estimators provides rejection frequencies that tend to be quite close to 0. As expected, when $c > 0$ the power increases. The normal approximation with the least-squares estimator implies higher rejection frequencies; however, the difference in power with the modified pairs bootstrap is always smaller than 0·15.

Remark 1. In equations (5) and (6), we recentre the bootstrap penalization term with respect to $\bar{\beta}_{n,i} I(|\bar{\beta}_{n,i}| \leqslant a_n)$ and $\bar{\beta}_{n,i} I(\check{\beta}_{n,i} = 0)$, respectively. These optimization problems can be solved using conventional algorithms developed for standard lasso estimators. Indeed, upon making the substitutions $\dot{\gamma}_i = u_i - \bar{\beta}_{n,i} I(|\bar{\beta}_{n,i}| \leqslant a_n)$ and $\ddot{\gamma}_i = u_i - \bar{\beta}_{n,i} I(\check{\beta}_{n,i} = 0)$, we obtain the standard lasso estimation criterion.

Remark 2. The accuracy of the lasso depends heavily on the selection of λ_n. Furthermore, in the definition of the modified pairs bootstrap we also have to select the sequence a_n. Using the results in

Theorem 1, we can develop a data-driven method for the selection of these tuning parameters in the spirit of Hall et al. (2009) and Chatterjee & Lahiri (2011). Its rationale is to select tuning parameters that minimize the estimated mean squared error of $\hat{\beta}_n$. For $\lambda_n \in [0, +\infty)$ and $a_n \in (0, +\infty)$, let $\dot{\beta}_n^* = \dot{\beta}_n^*(\lambda_n, a_n)$ denote the bootstrap modified lasso estimator. Using Theorem 1, we can estimate the mean squared error $E(\|\hat{\beta}_n - \beta\|^2)$ by

$$\phi(\lambda_n, a_n) = E^* \left\{ \|\dot{\beta}_n^*(\lambda_n, a_n) - \bar{\beta}_n\|^2 \right\}. \qquad (7)$$

Finally, we choose the optimal values $(\hat{\lambda}_n, \hat{a}_n)$ that minimize the estimated mean squared error (7). Using similar arguments, we can apply this procedure to select the tuning parameter λ_n for the adaptive lasso estimator $\check{\beta}_n$.

Remark 3. In our analysis, we have considered settings where the regression parameter β is fixed and the sample size n is large. The modified pairs bootstrap recentres the bootstrap statistic with respect to $\bar{\beta}_n$ and consequently relies on the consistency of $\bar{\beta}_n$. By using results of Huber & Ronchetti (2009) and imposing appropriate regularity conditions, it may be possible to extend the findings in this paper to settings where the dimension $p < n$ of the regression parameter β is allowed to depend on the sample size n. On the other hand, it remains unclear how to extend the definition of modified pairs bootstrap procedures to more general high-dimensional settings where p may be larger than n.

Acknowledgement

The author would like to thank the editor, the associate editor and two referees for helpful comments which have led to an improved version of the paper.

Supplementary material

Supplementary material available at *Biometrika* online includes proofs of the theoretical results, additional simulation results, and examples to clarify the reasons for failure of the naive pairs bootstrap for lasso estimators.

References

Bühlmann, P. & van de Geer, S. (2011). *Statistics for High Dimensional Data: Methods, Theory and Applications*. Heidelberg: Springer.

Chatterjee, A. & Lahiri, S. N. (2010). Asymptotic properties of the residual bootstrap for lasso estimators. *Proc. Am. Math. Soc.* **138**, 4497–509.

Chatterjee, A. & Lahiri, S. N. (2011). Bootstrapping lasso estimators. *J. Am. Statist. Assoc.* **106**, 608–25.

Chatterjee, A. & Lahiri, S. N. (2013). Rates of convergence of the adaptive lasso estimators to the oracle distribution and higher order refinements by the bootstrap. *Ann. Statist.* **41**, 1232–59.

Freedman, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9**, 1218–28.

Hall, P., Lee, E. R. & Park, B. U. (2009). Bootstrap-based penalty choice for the lasso achieving oracle performance. *Statist. Sinica* **19**, 449–71.

Huber, P. J. & Ronchetti, E. M. (2009) *Robust Statistics*. New York: Wiley.

Knight, K. & Fu, W. (2000). Asymptotics for lasso-type estimators. *Ann. Statist.* **28**, 1356–78.

Lockhart, R., Taylor, J., Tibshirani, R. J. & Tibshirani, R. (2014). A significant test for the lasso. *Ann. Statist.* **42**, 413–68.

Minnier, J., Tian, L. & Cai, T. (2011). A perturbation method for inference on regularized regression estimates. *J. Am. Statist. Assoc.* **106**, 1371–82.

Tibshirani, R. J. (1996). Regression analysis and selection via the lasso. *J. R. Statist. Soc.* B **21**, 267–88.

van de Geer, S., Bühlmann, P., Ritov, Y. & Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42**, 1166–202.

Wagener, J. & Dette, H. (2012). Bridge estimators and the adaptive lasso under heteroscedasticity. *Math. Meth. Statist.* **21**, 109–26.

Zhang, C. H. & Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Statist. Soc.* B **76**, 217–42.

Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Am. Statist. Assoc.* **101**, 1418–29.

[*Received June* 2014. *Revised June* 2015]